

RECEIVED

#22

MAY 13 2003
TECH CENTER 1600/2500

1600

RAW SEQUENCE LISTING

DATE: 05/05/2003

PATENT APPLICATION: US/09/858,332C

TIME: 09:43:03

Input Set : N:\Crf4\04182003\I858332B.raw

Output Set: N:\CRF4\05052003\I858332C.raw

1 <110> APPLICANT: Tchaga, Grigory S.
 2 Jokhadze, George
 3 <120> TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for
 4 Using the Same
 5 <130> FILE REFERENCE: CLON-056CIP
 C--> 6 <140> CURRENT APPLICATION NUMBER: US/09/858,332C
 7 <141> CURRENT FILING DATE: 2003-04-15
 8 <150> PRIOR APPLICATION NUMBER: 09/404,017
 9 <151> PRIOR FILING DATE: 1999-09-23
 10 <150> PRIOR APPLICATION NUMBER: 60/101,867
 11 <151> PRIOR FILING DATE: 1998-09-25
 12 <160> NUMBER OF SEQ ID NOS: 20
 13 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 15 <210> SEQ ID NO: 1
 16 <211> LENGTH: 16
 17 <212> TYPE: PRT
 18 <213> ORGANISM: Artificial Sequence
 19 <220> FEATURE:
 20 <223> OTHER INFORMATION: affinity peptide
 21 <400> SEQUENCE: 1
 22 His Leu Ile His Asn Val His Lys Glu Glu His Ala His Ala His Asn
 23 1 5 10 15
 25 <210> SEQ ID NO: 2
 26 <211> LENGTH: 18
 27 <212> TYPE: PRT
 28 <213> ORGANISM: Artificial Sequence
 29 <220> FEATURE:
 30 <223> OTHER INFORMATION: affinity peptide
 31 <400> SEQUENCE: 2
 32 His Asp Asp His Asp Asp His Asp Asp His Asp Asp His Asp Asp His
 33 1 5 10 15
 34 Asp Asp
 36 <210> SEQ ID NO: 3
 37 <211> LENGTH: 16
 38 <212> TYPE: PRT
 39 <213> ORGANISM: Artificial Sequence
 40 <220> FEATURE:
 41 <223> OTHER INFORMATION: affinity peptide
 42 <400> SEQUENCE: 3
 43 His Glu Glu His Glu Glu His Glu Glu His Glu Glu His
 44 1 5 10 15
 46 <210> SEQ ID NO: 4
 47 <211> LENGTH: 18

ENTERED

RAW SEQUENCE LISTING

DATE: 05/05/2003

PATENT APPLICATION: US/09/858,332C

TIME: 09:43:03

Input Set : N:\Crf4\04182003\I858332B.raw

Output Set: N:\CRF4\05052003\I858332C.raw

48 <212> TYPE: PRT
49 <213> ORGANISM: Artificial Sequence
50 <220> FEATURE:
51 <223> OTHER INFORMATION: affinity peptide
52 <400> SEQUENCE: 4
53 His Asp Glu His Asp Glu His Asp Glu His Asp Glu His
54 1 5 10 15
55 Asp Glu
57 <210> SEQ ID NO: 5
58 <211> LENGTH: 18
59 <212> TYPE: PRT
60 <213> ORGANISM: Artificial Sequence
61 <220> FEATURE:
62 <223> OTHER INFORMATION: affinity peptide
63 <400> SEQUENCE: 5
64 His Glu Asp His Glu Asp His Glu Asp His Glu Asp His
65 1 5 10 15
66 Glu Asp
68 <210> SEQ ID NO: 6
69 <211> LENGTH: 4
70 <212> TYPE: PRT
71 <213> ORGANISM: Artificial Sequence
72 <220> FEATURE:
73 <223> OTHER INFORMATION: enterokinase cleavage site
74 <400> SEQUENCE: 6
75 Ile Glu Gly Arg
76 1
78 <210> SEQ ID NO: 7
79 <211> LENGTH: 6
80 <212> TYPE: PRT
81 <213> ORGANISM: Artificial Sequence
82 <220> FEATURE:
83 <223> OTHER INFORMATION: a factor Xa cleavage site
84 <400> SEQUENCE: 7
85 Leu Val Pro Arg Gly Ser
86 1 5
88 <210> SEQ ID NO: 8
89 <211> LENGTH: 8
90 <212> TYPE: PRT
91 <213> ORGANISM: Artificial Sequence
92 <220> FEATURE:
93 <223> OTHER INFORMATION: a thrombin cleavage site
94 <400> SEQUENCE: 8
95 His Pro Phe His Leu Val Ile His
96 1 5
98 <210> SEQ ID NO: 9
99 <211> LENGTH: 10
100 <212> TYPE: PRT
101 <213> ORGANISM: Artificial Sequence

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/858,332C

DATE: 05/05/2003

TIME: 09:43:03

Input Set : N:\Crf4\04182003\I858332B.raw

Output Set: N:\CRF4\05052003\I858332C.raw

```

102 <220> FEATURE:
103 <223> OTHER INFORMATION: a renin cleavage site
104 <400> SEQUENCE: 9
105      Cys Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
106      1                5                10
108 <210> SEQ ID NO: 10
109 <211> LENGTH: 8
110 <212> TYPE: PRT
111 <213> ORGANISM: Artificial Sequence
112 <220> FEATURE:
113 <223> OTHER INFORMATION: an immunological tag
114 <400> SEQUENCE: 10
115      Asp Tyr Lys Asp Asp Asp Asp Lys
116      1                5
118 <210> SEQ ID NO: 11
119 <211> LENGTH: 11
120 <212> TYPE: PRT
121 <213> ORGANISM: Artificial Sequence
122 <220> FEATURE:
123 <223> OTHER INFORMATION: an immunological tag
124 <400> SEQUENCE: 11
125      Cys Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
126      1                5                10
128 <210> SEQ ID NO: 12
129 <211> LENGTH: 11
130 <212> TYPE: PRT
131 <213> ORGANISM: Artificial Sequence
132 <220> FEATURE:
133 <223> OTHER INFORMATION: an immunological tag
134 <400> SEQUENCE: 12
135      Cys Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
136      1                5                10
138 <210> SEQ ID NO: 13
139 <211> LENGTH: 3430
140 <212> TYPE: DNA
141 <213> ORGANISM: Artificial Sequence
142 <220> FEATURE:
143 <223> OTHER INFORMATION: DNA sequence of vector containing cDNA of
144      recombinant enterokinase
145 <400> SEQUENCE: 13
146      gacgaaaggg cctcgtgata cgcctat tttt tatagggttaa tgtcatgata ataatggttt 60
147      cttagacgtc aggtggcact tttcggggaa atgtgcgcgg aaccctatt tgtttat tttt 120
148      tctaaatata ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat 180
149      aatattgaaa aaggaagagt atgagtattc aacatttccg tgcgcgccctt attccctttt 240
150      ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 300
151      ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga 360
152      tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc 420
153      tatgtggcgc ggtattatcc cgtattgacg cggggcaaga gcaactcggg cgccgcatac 480
154      actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg 540

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/858,332C

DATE: 05/05/2003

TIME: 09:43:03

Input Set : N:\Crf4\04182003\I858332B.raw

Output Set: N:\CRF4\05052003\I858332C.raw

155	gcatgacagt	aagagaatta	tgcagtgtctg	ccataaccat	gagtgataac	actgcggcca	600
156	acttacttct	gacaacgatc	ggaggaccga	aggagctaac	cgcttttttg	cacaacatgg	660
157	gggatcatgt	aactcgcctt	gatcgttggg	aaccggagct	gaatgaagcc	ataccaaacg	720
158	acgagcgtga	caccacgatg	cctgtagcaa	tggcaacaac	gttgcgcaaa	ctattaactg	780
159	gcgaactact	tactctagct	ccccggcaac	aattaataga	ctggatggag	gcggataaag	840
160	ttgcaggacc	acttctgcgc	tcggcccttc	cggctggctg	gtttattgct	gataaatctg	900
161	gagccggtga	gcgtgggtct	cgcggtatca	ttgcagcact	ggggccagat	ggtaagccct	960
162	cccgatcatg	agttatctac	acgacgggga	gtcaggcaac	tatggatgaa	cgaaatagac	1020
163	agatcgctga	gataggtgcc	tactgatta	agcattggta	actgtcagac	caagtttact	1080
164	catatatact	ttagattgat	ttaaaacttc	atttttaatt	taaaaggatc	taggtgaaga	1140
165	tcctttttga	taatctcatg	acaaaaatcc	cttaacgtga	gttttcgttc	cactgagcgt	1200
166	cagaccccg	agaaaagatc	aaaggatctt	cttgagatcc	tttttttctg	cgcgtaatct	1260
167	gctgcttgca	aacaaaaaaa	ccaccgctac	cagcggtggt	ttgtttgccg	gatcaagagc	1320
168	taccaactct	ttttccgaag	gtaactggct	tcagcagagc	gcagatacca	aatactgtcc	1380
169	ttctagtgtg	gccgtagtta	ggccaccact	tcaagaactc	tgtagcaccg	cctacatacc	1440
170	tcgctctgct	aatcctgtta	ccagtggctg	ctgccagtgg	cgataagtgc	tgtcttaccg	1500
171	ggttggaactc	aagacgatag	ttaccggata	aggcgcagcg	gtcgggctga	acgggggggt	1560
172	cgtgcacaca	gccagccttg	gagcgaacga	cctacaccga	actgagatac	ctacagcgtg	1620
173	agctatgaga	aagcgccacg	cttcccgaag	ggagaaaggg	ggacaggtat	ccggtaagcg	1680
174	gbgcagggtc	ggaacaggag	agcgacgag	ggagcttcca	gggggaaacg	cctggatatct	1740
175	ttatagtcct	gtcgggtttc	gccacctctg	acttgagcgt	cgatttttgt	gatgctcgtc	1800
176	agggggggcg	agcctatgga	aaaacgccag	caacgcggcc	tttttacggt	tcctggcctt	1860
177	ttgctggcct	tttgetcaca	tgttctttcc	tgcgttatcc	cctgattctg	tggataaccg	1920
178	tattaccgcc	tttgagttag	ctgataccgc	tcgccgcagc	cgaacgaccg	agcgcagcga	1980
179	gtcagttagc	gaggaagcgg	aagagcgccc	aatacgcaaa	ccgcctctcc	ccgcgcgttg	2040
180	gccgattcat	taatgcagct	ggcacgacag	gtttcccgac	tggaaagcgg	gcagttagcg	2100
181	caacgcaatt	aatgtgagtt	agctcactca	ttaggcaccc	caggctttac	actttatgct	2160
182	tcgggctcgt	atgttggtg	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	2220
183	tgaccatgat	tacgccaagc	ttgaaggatc	atctcatcca	caatgtccac	aaagaggagc	2280
184	acgctcatgc	ccacaacaag	atcgatattg	tcggaggaag	tgactccaga	gaaggagcct	2340
185	ggccttggtg	cgttgctctg	tatttcgacg	atcaacaggt	ctgcggagct	tctctggtga	2400
186	gcagggattg	gctggtgtcg	gccgccact	gcgtgtacgg	gagaaatatg	gagccgtcta	2460
187	agtggaaagc	agtgtctaggc	ctgcataatg	catcaaactc	gacttctcct	cagatagaaa	2520
188	ctaggttgat	tgaccaaatt	gtcataaacc	cacactacaa	taaacggaga	aagaacaatg	2580
189	acattgccat	gatgcactct	gaaatgaaag	tgaactacac	agattatata	cagcctattt	2640
190	gtttaccaga	agaaaatcaa	gtttttcccc	caggaagaat	ttgttctatt	gctggctggg	2700
191	gggcacttat	atatcaaggt	tctactgcag	acgtactgca	agaagctgac	gttccccctc	2760
192	tatcaaata	gaaatgtcaa	caacagatgc	cagaatataa	cattacggaa	aatatggtgt	2820
193	gtgcaggcta	tgaagcagga	ggggtagatt	cttgtcaggg	ggattcaggc	ggaccactca	2880
194	tgtgccaaga	aaacaacaga	tggtcctcgg	ctggcgtagc	gtcatttgga	tatcaatgtg	2940
195	cactgcctaa	tcgcccaggg	gtgtatgcc	gggtcccaag	gttcacagag	tggatacaaa	3000
196	gttttctaca	tgagctcgta	attagctgag	aattcactgg	ccgtcgtttt	acaacgtcgt	3060
197	gagcctggga	aaaccctggc	gttaccacac	ttaatcgctc	tcgagcacat	ccccctttcg	3120
198	ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	ttcccaacag	ttgcgcagcc	3180
199	tgaatggcga	atggcgctg	atgcggtatt	ttctccttac	gcattctgtg	ggtattttcac	3240
200	accgcatatg	gtgcactctc	agtacaatct	gctctgatgc	cgcatagtta	agccagcccc	3300
201	gacacccgcc	aacacccgct	gacgcgccct	gacgggcttg	tctgctcccg	gcatecgctt	3360
202	acagacaagc	tgtgaccgtc	tcggggagct	gcattgtgtc	gaggttttca	ccgtcatcac	3420
203	cgaaacgcgc						3430

RAW SEQUENCE LISTING

DATE: 05/05/2003

PATENT APPLICATION: US/09/858,332C

TIME: 09:43:03

Input Set : N:\Crf4\04182003\I858332B.raw

Output Set: N:\CRF4\05052003\I858332C.raw

```

205 <210> SEQ ID NO: 14
206 <211> LENGTH: 279
207 <212> TYPE: PRT
208 <213> ORGANISM: Artificial Sequence
209 <220> FEATURE:
210 <223> OTHER INFORMATION: protein sequence of vector containing cDNA of
211 recombinant enterokinase
212 <400> SEQUENCE: 14
213 Met Thr Met Ile Thr Pro Ser Leu Lys Asp His Leu Ile His Asn Val
214 1 5 10 15
215 His Lys Glu Glu His Ala His Ala His Asn Lys Ile Asp Ile Val Gly
216 20 25 30
217 Gly Ser Asp Ser Arg Glu Gly Ala Trp Pro Trp Val Val Ala Leu Tyr
218 35 40 45
219 Phe Asp Asp Gln Gln Val Cys Gly Ala Ser Leu Val Ser Arg Asp Trp
220 50 55 60
221 Leu Val Ser Ala Ala His Cys Val Tyr Gly Arg Asn Met Glu Pro Ser
222 65 70 75 80
223 Lys Trp Lys Ala Val Leu Gly Leu His Met Ala Ser Asn Leu Thr Ser
224 85 90 95
225 Pro Gln Ile Glu Thr Arg Leu Ile Asp Gln Ile Val Ile Asn Pro His
226 100 105 110
227 Tyr Asn Lys Arg Arg Lys Asn Asn Asp Ile Ala Met Met His Leu Glu
228 115 120 125
229 Met Lys Val Asn Tyr Thr Asp Tyr Ile Gln Pro Ile Cys Leu Pro Glu
230 130 135 140
231 Glu Asn Gln Val Phe Pro Pro Gly Arg Ile Cys Ser Ile Ala Gly Trp
232 145 150 155 160
233 Gly Ala Leu Ile Tyr Gln Gly Ser Thr Ala Asp Val Leu Gln Glu Ala
234 165 170 175
235 Asp Val Pro Leu Leu Ser Asn Glu Lys Cys Gln Gln Gln Met Pro Glu
236 180 185 190
237 Tyr Asn Ile Thr Glu Asn Met Val Cys Ala Gly Tyr Glu Ala Gly Gly
238 195 200 205
239 Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Gln Glu
240 210 215 220
241 Asn Asn Arg Trp Leu Leu Ala Gly Val Thr Ser Phe Gly Tyr Gln Cys
242 225 230 235 240
243 Ala Leu Pro Asn Arg Pro Gly Val Tyr Ala Arg Val Pro Arg Phe Thr
244 245 250 255
245 Glu Trp Ile Gln Ser Phe Leu His Glu Leu Val Ile Ser Glu Phe Thr
246 260 265 270
247 Gly Arg Arg Phe Thr Thr Ser
248 275
250 <210> SEQ ID NO: 15
251 <211> LENGTH: 12
252 <212> TYPE: PRT
253 <213> ORGANISM: Artificial Sequence
254 <220> FEATURE:

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/858,332C

DATE: 05/05/2003

TIME: 09:43:04

Input Set : N:\Crf4\04182003\I858332B.raw

Output Set: N:\CRF4\05052003\I858332C.raw

L:6 M:270 C: Current Application Number differs, Wrong Format